Predicting the past with GIS at Indiana Dunes National Lakeshore

By Amanda Renner

Abstract

The National Park Service's Midwest Archeological Center is the repository for cultural resource site and survey location data collected for nearly five decades at more than 50 parks in the Midwest Region. As more and better data are accumulated, focus is now shifting toward integrating these large data sets to address "big picture" issues facing park resource managers. To that end, an ongoing multiyear archeological inventory at Indiana Dunes National Lakeshore included the creation of a GIS-based archeological site predictive model. The model integrates data collected from small compliance projects to large parkwide inventories, and identifies patterns in the environmental characteristics of site locations. This article outlines the protocol used to integrate multiple data sets and addresses questions regarding prehistoric land use as well as the nature of previous sampling, providing a point from which to continue scientific inquiry.

Key words

archeology, Geographic Information Systems (GIS), Great Lakes, Indiana Dunes National Lakeshore, predictive modeling

INDIANA DUNES NATIONAL LAKEshore was established in 1966 "in order to preserve for the educational, inspirational, and recreational use of the public . . . the Indiana dunes and other areas of scenic, scientific, and historic interest and recreational value" (16 U.S.C. 4604). Located along the southern shoreline of Lake Michigan, this 15-mile-long (24 km) park preserves roughly 15,000 acres (6,070 ha) of sand dunes, oak savannas, swamps, bogs, marshes, and forests (fig. 1). This tightly packed set of ecosystems harbors biological diversity that is among the highest per unit area in the National Park System, and is associated with a series of dunes that began to form as glaciers started receding from the area nearly 14,000 years ago, according to the Indiana Dunes National Lakeshore Web site (NPS 2013). Glacial retreat, in addition to wind and water action, resulted in spatially distinct dune systems, dating to the late Pleistocene, mid-Holocene, and recent past. The national lakeshore is notable for these chronologically distinct systems, and the study of ecology is even linked to this place because of the opportunity to observe the correlation of plant communities with landform age.

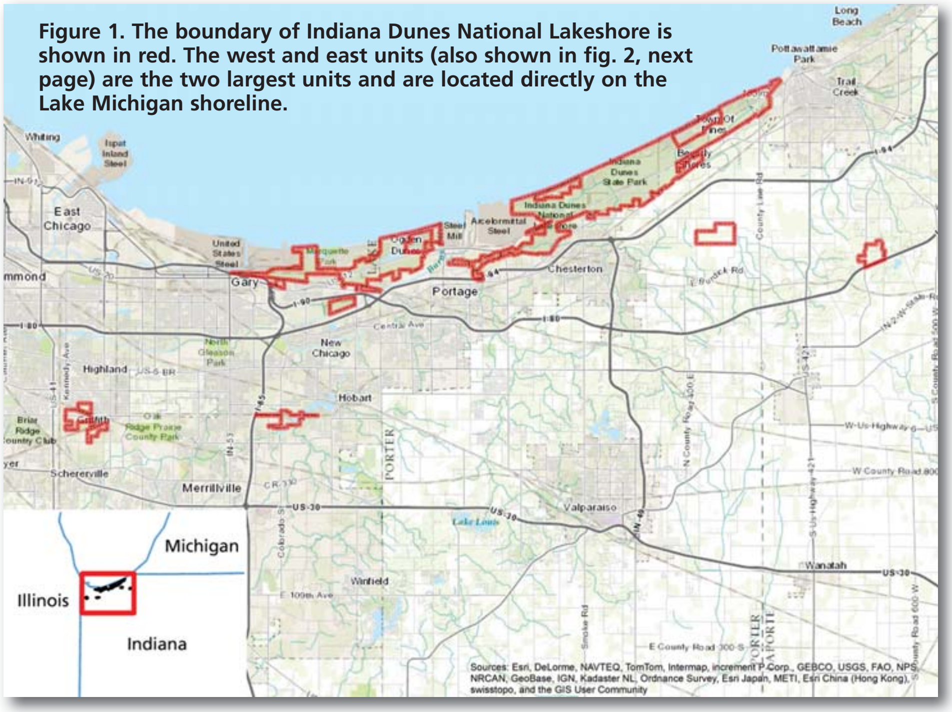

Figure 1. The boundary of Indiana Dunes National Lakeshore is shown in red. The west and east units (also shown in fig. 2, next page) are the two largest units and are located directly on the Lake Michigan shoreline.

The Midwest Archeological Center has fostered a relationship with Indiana Dunes National Lakeshore spanning four decades. Nearly 70 archeological projects have been conducted in that time, resulting in the inventory of some 1,000 acres (405 ha), or 7% of the park. Around 239 archeological sites have been recorded in the park (NPS 2012, Archeological Sites Management Information System), extending archeological evidence of human use of the southern Lake Michigan shoreline to the Late Paleo-Indian period approximately 10,000–8,500 years ago (Bringelson and Sturdevant 2007). While some of this information was acquired as part of research-oriented work representing large-tract sampling, most of it was gathered during small-scale projects for legal compliance purposes. In accordance with Section 106 of the National Historic Preservation Act of 1966, federal agencies must determine if their actions are likely to have an effect on historical properties, in this instance an archeological site. For example, shovel-testing of expiring lease properties is a simple method used to determine the presence of archeological sites, and is conducted as part of park planning. Regardless of the quantity of archeological work conducted here, the sample is limited in size and scope. This makes it difficult to draw conclusions about human land use along the lakeshore during the Holocene. Questions also linger regarding the apparent lack of occupation of certain areas within the park before the arrival of Europeans.

Answers sought

The Midwest Archeological Center is taking advantage of this large yet diffuse data set as part of a multiyear archeological inventory of the park. We have incorporated data from past projects with environmental data sets acquired from county, state, and federal government sources in order to create a GIS-based archeological site location predictive model. We obtained measurements for slope, elevation, and distance to water for archeological site and nonsite locations. We then conducted a Multivariate Logistic Regression (MLR) analysis to detect patterns in the environmental characteristics of prehistoric site locations and thus identify landforms with potential for past human use. Preliminary results indicated that archeological sites correlated with the sand dune and ridge formations, as researchers had previously assumed. During the 2012 field season, we inventoried and tested high-probability areas identified by the model and have confirmed seven new archeological sites.

The science of modeling

Archeological site location modeling typically compares the environmental characteristics of site and nonsite locations to predict where other sites are likely to be located in areas that have not been surveyed (Kvamme 1985, 1990, 1992). Typical environmental variables found to influence site location selection include proximity to water, slope, aspect, elevation, soil type, and vegetation type. We selected four variables to create the Indiana Dunes model: elevation, slope, soil type, and proximity to water other than the lake. Using GIS we measured each variable at archeological site and nonsite locations within the project area. Slope and elevation data were obtained from a 5-meter-resolution Digital Elevation Model created from contours derived from elevation data acquired by NPS staff from county and state sources. Distance to the nearest water source was obtained from the U.S. Environmental Protection Agency National Hydrography Data Set and from the USDA Geospatial Data Gateway (USGS 2012). Soils data were acquired from the Natural Resource Conservation Survey Soil Data Mart (NRCS 2012).

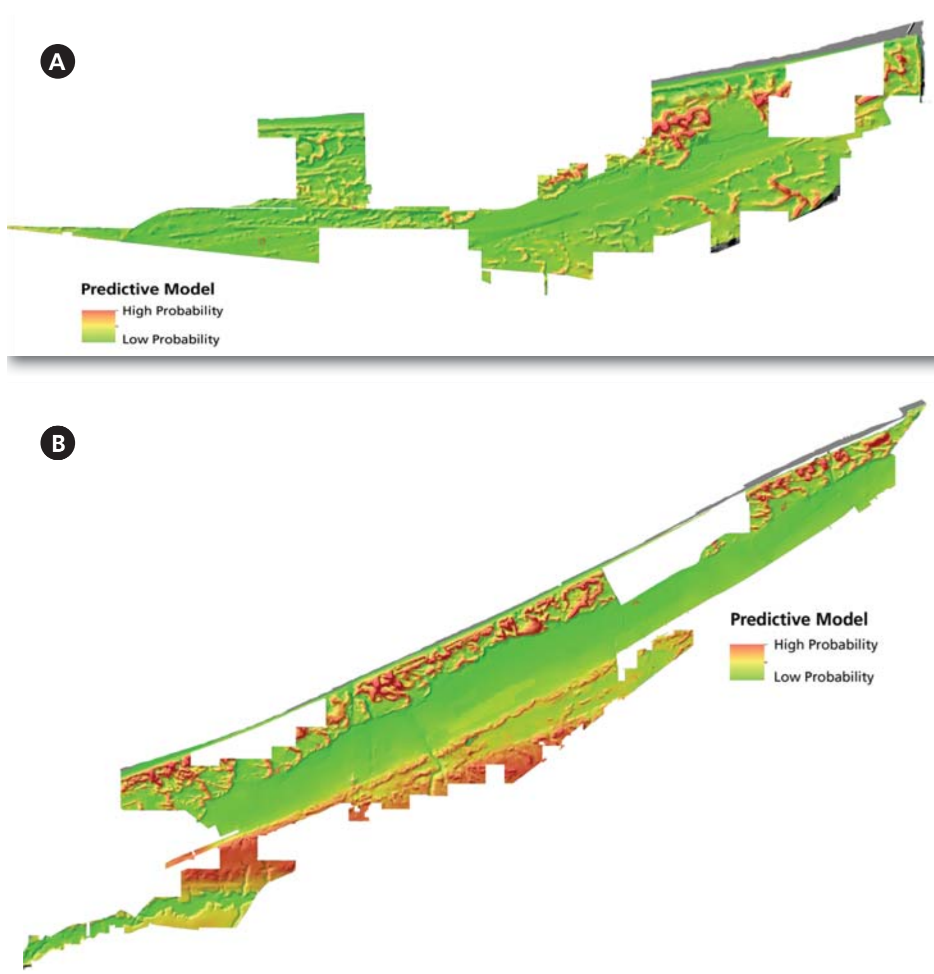

Figure 2. Archeological site predictive surface raster, Indiana Dunes National Lakeshore. Raster A (top) shows areas of high probability for archeological site locations (red), which are associated mainly with recent dune formations along the shoreline, and low-probability (green) lowland areas in the western unit. Raster B (bottom) shows areas of high probability for archeological site locations (red) associated with recent and older dune systems dating back to the late Pleistocene further inland and low-probability (green) lowland areas in the eastern unit.

We conducted a Multivariate Logistic Regression (MLR) analysis to determine how the combination of variables influenced prehistoric site selection in the area. We found three of the variables to have a significant influence on site location: elevation, slope, and soil type. The model correctly predicted 74% of nonsite locations and 62% of site locations. After performing this analysis, we used the resulting formula (next page) to create archeologi-

Typical environmental variables found to influence site location selection include proximity to water, slope, aspect, elevation, soil type, and vegetation type.

cal site predictive surface raster images as shown in figs. 2A and 2B.

The end result of the MLR analysis is a *Z* "score" representing the logarithmic odds of a positive response for the dependent variable, in this case archeological site presence:

$$Z = B_0 + B_{1X1} + B_{2X2} + B_{3X3} + B_{4X4}$$

In the formula, B_0 is the "constant" or value of *Z* when the independent variables are zero. B_1, B_2, and B_3 are the "regression coefficients" of the independent variables x1, x2, and x3, respectively.

We used the regression coefficient for each variable to weight each corresponding raster data layer in the GIS. This is accomplished using the raster calculator via the "Spatial Analyst" tool in ArcGIS 10.1. For example, using the raster calculator tool we multiplied each cell value in the slope data layer by the regression coefficient for slope, and we repeated this process for each variable. The layers were then added together with the value of the constant (B_0) to create a GIS raster data layer representing the *Z* score for each location on the landscape. In this layer each cell value is the *Z* score for that location. We then converted this "score" to a probability using the following equation:

$$1 / (1 + \text{exponent}^{(-\text{score})})$$

(Pampel 2000). This converts the *Z* score to a probability on a scale of 0–1, which in turn creates the predictive surface raster layer with cell values that represent the probability that the location will have an archeological site. Cell values range from 0 (low probability) to 1 (high probability).

Conclusion

The predictive surface raster model helped guide our selection of archeological inventory areas for the 2012 field season. For 2013 we also relied on the model but first fed results of the 2012 season back into it to help refine its predictive abilities. We will continue to follow this pattern as more data are collected. We are hopeful that this approach will expand our understanding of the archeological record at Indiana Dunes, including how well the current sample of sites represents the total record. Additionally, with the development of this method archeologists are better able to support planners, resource managers, and interpreters at the park. Ultimately, these methods will lead to further knowledge of the ways people used wetlands, dune ridges, and other key aspects of the Indiana Dunes ecosystem over time.

References

Bringelson, D., and J. T. Sturdevant. 2007. An archeological overview and assessment of Indiana Dunes National Lakeshore, Indiana. Midwest Archeological Center Technical Report 97. National Park Service, Lincoln, Nebraska, USA.

Kvamme, K. 1985. Determining empirical relationships between the natural environment and prehistoric site locations: A hunter-gatherer example. Pages 208–238 *in* C. Carr, editor. For concordance in archaeological analysis. Westport Publishers, Kansas City, Missouri, USA.

———. 1990. One-sample tests in regional archaeological analysis: New possibilities through computer technology. American Antiquity 55(2):367–381.

———. 1992. A predictive site location model on the high plains: An example with an independent test. Plains Anthropologist 37(138):19–40.

National Park Service (NPS). 2012. Archeological Sites Management Information System. https://64.241.25.40/Login/. Accessed April 2012.

———. 2013. Indiana Dunes National Lakeshore Web site, http://www.nps.gov/indu/naturescience/index.htm. Accessed April 2013.

Natural Resource Conservation Service (NRCS). 2012. Porter and Lake County soil data set. Soil Data Mart Web site, http://soildatamart.nrcs.usda.gov/. Accessed April.

Pampel, F. C. 2000. Logistic regression: A primer. Quantitative applications in the social sciences. Sage Publications, Newbury Park, California, USA.

United States Geologic Survey (USGS). 2012. National Hydrography Data Set. USDA Geospatial Data Gateway Web site, http://datagateway.nrcs.usda.gov/GDGHome.aspx. Accessed April 2012.

About the author

Amanda Renner *(amanda_renner@nps.gov) is an archeologist with the Midwest Archeological Center in Lincoln, Nebraska. She received her BA in anthropology from the University of Nebraska–Lincoln in 2005 and MA in anthropology from the University of Wyoming in 2009. She recently completed her graduate certificate in geographic information science from the University of Nebraska–Lincoln. Her interests include applications of GIS to archeology, predictive modeling, and Great Plains/Rocky Mountains archeology.*